**Assignment 1 – Dissimilarity Metrics**

Consider the tables of data from health records shown below.

The first table gives general health information for 10 subjects.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Age (yrs)** | **Gender**  **(M/F)** | **Weight**  **(lbs)** | **Height**  **(in)** | **Blood**  **Pressure**  **(systolic)** | **Blood Pressure**  **(diastolic)** | **Heart Rate**  **(bpm)** | **Smoker**  **(Y/N)** | **Diabetic**  **(Y/N)** | **Family History of Heart Disease**  **(Y/N)** | **Family**  **History**  **Of**  **Cancer**  **(Y/N)** | **Family**  **History**  **Of**  **Diabetes**  **(Y/N)** | **Family**  **History**  **Of**  **Stroke**  **(Y/N)** |
| **1** | **10** | **M** | **68** | **58** | **125** | **82** | **80** | **N** | **Y** | **Y** | **N** | **Y** | **N** |
| **2** | **45** | **M** | **175** | **73** | **121** | **78** | **68** | **N** | **N** | **N** | **Y** | **N** | **N** |
| **3** | **35** | **F** | **156** | **63** | **115** | **70** | **82** | **N** | **N** | **N** | **N** | **N** | **N** |
| **4** | **18** | **M** | **210** | **75** | **123** | **81** | **90** | **N** | **N** | **N** | **Y** | **N** | **N** |
| **5** | **51** | **F** | **145** | **64** | **135** | **83** | **85** | **N** | **N** | **Y** | **N** | **Y** | **N** |
| **6** | **63** | **F** | **131** | **66** | **155** | **91** | **65** | **Y** | **N** | **N** | **N** | **N** | **Y** |
| **7** | **27** | **M** | **182** | **71** | **142** | **85** | **73** | **Y** | **N** | **N** | **N** | **N** | **N** |
| **8** | **28** | **F** | **125** | **60** | **118** | **75** | **70** | **N** | **N** | **N** | **N** | **N** | **N** |
| **9** | **75** | **M** | **219** | **72** | **151** | **88** | **91** | **Y** | **Y** | **N** | **Y** | **N** | **N** |
| **10** | **77** | **M** | **158** | **68** | **140** | **92** | **75** | **N** | **N** | **N** | **Y** | **N** | **N** |

**Table 1.** *General Health Records*

For the same subjects, a general health questionnaire was given, with the results shown below:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **General Feeling**  **(unwell-well)**  **(1 – 5)** | **Sleep**  **(tired – energetic)**  **(1 – 5)** | **Colds per Year**  **(few – many)**  **(1 – 5)** | **Exercise**  **(never, seldom, sometimes, often)** | **Doctor’s Visits**  **(never, seldom, sometimes, often)** | **Lifestyle**  **(sporty, busy, stressful, sedentary)** | **Eyesight**  **(farsighted, nearsighted,**  **both, neither)** | **Eating Habits**  **(vegan, vegetarian, carnivorous)** | **Primary Alcohol Consumption**  **(beer, wine, hard liquor, none)** |
| **1** | **2** | **1** | **3** | **seldom** | **often** | **sedentary** | **both** | **carnivorous** | **none** |
| **2** | **4** | **4** | **2** | **sometimes** | **seldom** | **busy** | **neither** | **carnivorous** | **beer** |
| **3** | **5** | **4** | **2** | **often** | **seldom** | **stressful** | **neither** | **vegetarian** | **wine** |
| **4** | **5** | **5** | **2** | **often** | **never** | **sporty** | **neither** | **carnivorous** | **beer** |
| **5** | **3** | **3** | **3** | **sometimes** | **seldom** | **busy** | **nearsighted** | **carnivorous** | **wine** |
| **6** | **3** | **4** | **5** | **sometimes** | **sometimes** | **stressful** | **farsighted** | **vegetarian** | **hard liquor** |
| **7** | **2** | **3** | **2** | **never** | **sometimes** | **sedentary** | **neither** | **carnivorous** | **beer** |
| **8** | **4** | **5** | **1** | **often** | **seldom** | **sporty** | **neither** | **vegan** | **none** |
| **9** | **2** | **2** | **5** | **seldom** | **often** | **sedentary** | **farsighted** | **carnivorous** | **hard liquor** |
| **10** | **3** | **3** | **4** | **sometimes** | **sometimes** | **sedentary** | **both** | **carnivorous** | **beer** |

**Table 2.** *Health Questionnaire*

**Q1.** From **Table 1**, determine dissimilarity matrices for the **numeric data** using the following metrics:

(a) Euclidean

(b) Manhattan

(c) Supremum

**Q2.** From **Table 1**, determine dissimilarity matrices for the **binary data**:

(a) Assuming that all of the attributes are symmetric (Y and N are equally important)

(b) Assuming that all of the attributes are asymmetric (where we consider Y to be unimportant)

**Q3.** From **Table 1**, determine a dissimilarity matrix for the **entire table**, (assuming that the binary attributes are symmetric and that we used a Manhattan distance for numeric data).

**Q4.** From **Table 2**, determine a dissimilarity matrix for **ordinal data** (assume that the first 5 columns are ordinal).

**Q5.** From **Table 2**, determine a dissimilarity matrix for **nominal data** (assume that the last 5 columns are nominal).

**Q6.** From **Table 2**, determine a dissimilarity matrix for the **entire table** (under the same assumptions as for Q4 and Q5.

**Q7.** For the combination of **both tables**, determine a dissimilarity matrix (under the assumptions given in Q3, Q4 and Q5).